Amendments to the Specification:

Please add disclosure following paragraph [0120] with the following paragraphs:

One such example as described in Provisional Application Serial No. 60/398,958, and attorney docket No. 3508.1 (Provisional Application Serial No. 60/422,220) each incorporated by reference above includes a process of determining relative transcript concentrations from hybridization intensity data from experiments with probe arrays and gene structure information using model fitting.

In the presently described example, a transcript may include several gene features which refer to the sequences extracted from different splicing variants of the genes. A gene feature can be either exon, intron, or junction (exon-exon junction, exon-intron junction, intron-exon junction). An exon feature can be partitioned further depending on whether the exon is cassette exon or exons overlapping with others. Probes targeted to these features could be mapped to each of the transcript.

Gene structure includes all the transcripts of each gene and feature composition for each transcript. For example, a gene could have two transcripts A and B. Transcript A included 3 of the 5 features while transcript B had 4 features. The relationship between features and transcripts could be represented by matrix with values of 1s or 0s described as follows:

Let G be an m by n matrix, where m is the number of transcripts while n represents the number of features for a gene. The column $F^{(l)}$ denotes feature l, and $T_{j,k}$ denotes k^{th} transcript, it is also used to denote the concentration of k^{th} transcript in

experiment j. $g_{k,l}$ is the element of this matrix for k^{th} transcript and l^{th} feature, its value is either 1 or 0.

The matrix could be written using the following equation, where value $X_j^{(l)}$ denotes the concentration measured by l^{th} feature in experiment j. $X_j^{(l)}$ could be written as:

$$\forall k, X_j^{(l)} = \sum_{k=1}^{m} g_{k,l} T_{k,j} \qquad (1)$$

Equation (1) therefore represents the gene structure.

Continuing with the present example, to model the data multiple probes may be employed to represent each feature. In the present example, these probes typically have different properties, however they measure the same concentration of a given transcript feature.

A simple model may be adopted to express the relationship between probes properties, concentrations and intensity measurements:

$$y_{ij} = a_i x_j + \varepsilon_{ij} \qquad (2)$$

$$y_{ij} = a_i x_j + b_i + \varepsilon_{ij} \qquad (3)$$

In the above equations, a_i represents the affinity term for i^{th} probe (which is arbitrarily assigned), b_i represents the background index for i^{th} probe. These terms are

probe-dependent. Also, x_j represents the relative concentration of the feature in j^{th} experiment, and ε_{ij} denotes the error term. The error term included all factors not explained by the other terms, usually it is assumed to be normal distribution with mean 0 and variance σ^2. Formally, this could be written as $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The above formulas could be rewritten as follows for the *f(k)th* feature of a given gene:

$$y_{ij}^{f(k)} = a_i^{f(k)} x_j^{f(k)} + \varepsilon_{ij} \qquad (4)$$

$$y_{ij}^{f(k)} = a_i^{f(k)} x_j^{f(k)} + b_i^{f(k)} + \varepsilon_{ij} \qquad (5)$$

Combining these equations with equation (1), we have for feature *f(k)* of a gene:

$$y_{ij}^{f(k)} = a_i^{f(k)} \sum_{k=1}^{m} g_{k,f(k)} T_{k,j} + \varepsilon_{ij} \qquad (6)$$

$$y_{ij}^{f(k)} = a_i^{f(k)} \sum_{k=1}^{m} g_{k,f(k)} T_{k,j} + b_i^{f(k)} + \varepsilon_{ij} \qquad (7)$$

Differences between the predicated and observed intensity for each probe is thus minimized. A loss function may be required for penalizing errors in predication. Many types of loss functions may be used for the same purpose, such as squared error loss function, absolute difference loss function. In the present example, the squared error loss function is applied to the model.

To minimize the squared difference between predicated and observed intensity value for all the probes of each gene (a set of features), the equations could be written as:

$$f(\overrightarrow{A},\overrightarrow{T}) = \sum_{k=1}^{nf}\sum_{j=1}^{ne}\sum_{i=1}^{np}(y_{ij}^{(k)} - a_i^{(k)}x_j^{(k)})^2 = \sum_{k=1}^{nf}\sum_{j=1}^{ne}\sum_{i=1}^{np}(y_{ij}^{(k)} - a_i^{(k)}(\sum_{k=1}^{m} g_{k,f(k)}T_{k,j}))^2 \quad (8)$$

$$f(\overrightarrow{A},\overrightarrow{T}) = \sum_{k=1}^{nf}\sum_{j=1}^{ne}\sum_{i=1}^{np}(y_{ij}^{(k)} - a_i^{(k)}x_j^{(k)} - b_i^{(k)})^2 = \sum_{k=1}^{nf}\sum_{j=1}^{ne}\sum_{i=1}^{np}(y_{ij}^{(k)} - a_i^{(k)}(\sum_{k=1}^{m} g_{k,f(k)}T_{k,j}) - b_i^{(k)})^2 \quad (9)$$

To minimize $f(\overrightarrow{A},\overrightarrow{T})$, some constraints or penalty terms are needed in order to solve the equations. The following constraints may be added:

(10) $\sum_{i=1}^{np}(a_i^{(k)})^2 = constant$

(11) $a_i^{(k)} > 0$

(12) $T_{k,j}^{(k)} > 0$

Alternatively, the following penalty terms could be added to equations (7) and (8),

$$\gamma\sum_{k=1}^{nf}\sum_{i=1}^{np}(a_i^{(k)})^2$$

Maximum likelihood estimation is used. The solution may obtained by iteratively solving different sets of the parameters until convergence, yielding the relative concentration of each variant and the relative affinity term of each probe.